DOCUMENT RESUME

ED 373 096 TM 021 987

AUTHOR Kunce, Charles S.; Arbet, Scott E.
TITLE Issues of Candidate Perception in a Performance Test for Lawyers.
PUB DATE Apr 94
NOTE 37p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 5-7, 1994).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Attitudes; Demography; *Languages; *Lawyers; *Licensing Examinations (Professions); Pilot Projects; Questionnaires; *Test Construction; Test Format; Test Length; Test Validity
IDENTIFIERS *Bar Examinations; National Conference of Bar Examiners; *Performance Based Evaluation

ABSTRACT

The National Conference of Bar Examiners commissioned American College Testing, Inc., to help them in the development and evaluation of a performance test for use in bar admissions decisions. Because it was recognized that candidate perceptions would provide valuable information, a candidate-perception questionnaire was developed to be administered to candidates in the pilot study of the performance test. Demographic and historical information was requested, as well as information about candidates' perceptions of the test. In this study, perceptions of candidates who took the 3-hour test version were compared with those of candidates who took the 90-minute version. The number of candidates completing questionnaire items varied, but the sample was approximately 400 for the 3-hour test and 1,280 for the 90-minute test. Candidates' perceptions about the tests varied widely, but two conclusions are unmistakable: Candidates on the whole accept the face validity of the performance test, and candidates taking either test, regardless of length, felt that they did not have enough time to complete the performance tasks. Eighteen graphs present study findings. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 373 096

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
CHARLES S. KUNCE
SCOTT E. ARBET

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Issues of Candidate Perception in a Performance Test for Lawyers

Charles S. Kunce
Scott E. Arbet

American College Testing

April 1994

Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, in Symposium P3, Performance Assessment: Evaluating Competing Perspectives on Use for High Stakes Testing.

BEST COPY AVAILABLE

TM021987

Overview

The National Conference of Bar Examiners (NCBE) commisioned ACT to assist them in the development and evaluation of a performance test (PT) for use in Bar admissions decisions. Early in the planning of the PT, the NCBE and ACT recognized that candidate perceptions could provide valuable information in the development of the program. Therefore it was determined that a questionnaire would be administered to each candidate in the pilot study immediately upon completion of the PT. The questionnaires were developed by ACT in consultation with NCBE, and were administered by each jurisdiction's administrative staff. ACT tabulated and analyzed the data.

The items in the questionnaire were designed to gather meaningful demographic and historical information about each candidate plus specific opinions about various aspects of the PT. It was hoped that the analysis would provide useful information for improving the examinations, as well as indicate their current level of acceptability to candidates. This study was designed to address the following questions:

1. Do candidates perceive the performance test as a valid, accurate, and important measure of legal skills?
2. Did the candidates have adequate time and materials to complete the PT?
3. Were there differences in candidate perceptions between the 3-hour and 90-minute conditions?
4. Were there any meaningful correlations between the PT performance and questionnaire items?

The candidates could have been grouped in a number of different ways. For the purposes of this study, we decided to look at candidates who took the 3-hour PT in relation to those who took the 90-minute PT, because the comparability of these two groups was relevant to the question of whether the PT could be offered in the shorter format. Summary statistics are therefore provided for candidate responses based on total group, 3-hour condition, and 90-minute condition. Correlations of performance on the PT with a number of demographic and opinion questions were also investigated. This "data exploration" was intended to provide additional clues to the reasons for certain opinions, as well as a better understanding of any demographic correlates.

Naturally, candidate perceptions varied widely, but two conclusions are unmistakable: (1) Candidates as a whole accept the face validity of the PT, and (2) Candidates in both the 3-hour and the 90-minute conditions felt that they did not have enough time to complete the PT tasks.

Methodology

The questionnaire was designed to address three general areas of inquiry:

1. Acceptability of the concept. These questions concern perceptions of the relative value of the performance test, including its accuracy, importance, and apparent validity. Examples of this are items 12, 13, and 17.
2. Perceived flaws in the design. These questions concern opinions on the sufficiency of time allowed, the adequacy of the materials provided, etc. Examples of this are items 16, 18, and 19.
3. Performance and opinion correlates. These include a number of demographic and background variables such as age, gender, undergraduate major, language history, and experience. Items 1 through 11 are of this type.

Copies of the questionnaire were distributed to all PT testing sites in three pilot jurisdictions. After candidates completed all portions of the examinations, they were asked to complete questionnaires before leaving. Candidates were asked to provide their state-assigned applicant number. However, they were assured that individual responses to the questionnaire would be confidential and would not affect the grading of the examinations. Responses were marked directly in the questionnaire booklet. Data entry was carried out at ACT. There was no time limit on the questionnaire.

Data Preparation

First, data were sorted by 90-minute vs. 3-hour conditions. Total sample consists of the total of the 90-minute and 3-hour conditions. Examinees who were provided special testing arrangements (such as extended time, large print, or typing) were not included in these tabulations.

Except as described below, missing data (usually the result of people skipping questions) were excluded from the analyses of individual items. For this reason, the total number of responses (N) will vary from question to question. Means, standard deviations (SDs), and ranges are calculated on the Ns shown.

Additional special data preparation was required for four of the items:

Item 2 (Birth date): First, an index of age was calculated by subtracting the birth year from 93. Then, resulting values below 18 were omitted because of probable invalidity.

Item 5 (Reported LSAT score): LSAT scores from June 1982 to February 1991 were on a score scale resulting in 2-digit scores. The large majority of candidates reported 2-digit scores between 10 and 48, which could be actual scores on that score scale. Because this score scale and the more recent 3-digit score scale are not combinable for these analyses, only scores in the range of 10 to 48 were tabulated. The Ns shown are of the actual number of cases used.

Item 8 (Semester credits earned in certain types of courses): For the two variables under this item, it appeared that missing data would most logically represent zero credits earned. Therefore, in the summary statistics for item 8, omits (blanks) are entered as zeros.

Item 17 (Relative weight for each test): In many cases, candidates apparently skipped response areas rather than entering zeros. Given the context of the question, it was judged appropriate to enter blanks as zeros as we tabulated the data.

Data Tabulation

After the data were prepared, examinee responses were tabulated for each question. The total number of responses tabulated (N) is shown, as are the means and standard deviations (SD) of the distributions of responses. Where appropriate, the range of responses is also shown. For ease of reference, each questionnaire item is shown followed by the summary statistics. The results shown allow direct comparison of the 3-hour, 90-minute, and total groups' demographic and opinion information.

Item 22 at the end of the questionnaire invited free-form comments from examinees. Most candidates wrote at least a half page. ACT staff analyzed these responses by reading every one and tallying the themes that emerged. A summary of these themes is provided.

Relationship Analyses

The presence of demographic, attitudinal, and performance information invited some "data exploration." Of particular interest was whether the demographic or opinion variables were related to actual total score on the performance test. Pearson product-moment correlations were calculated between total scaled scores[1] on the PT and

[1]Raw scores for each task were adjusted to a common scale to account for possible differences in difficulty of tasks. Because tasks had been administered to randomly equivalent groups of examinees, it was technically feasible to scale the scores for each task to have a mean of 3 and a standard deviation of 1. The scaling process

many of the variables in the questionnaire. Chi-square statistics were used to test the null hypotheses that the correlations were equal to zero in all cases. Alpha levels were set at 0.01.

Results

Response Tabulations

Each of the questions is reproduced below, followed by summary statistics for total group, 3-hour, and 90-minute conditions. Graphs of results are also provided for items 12, 13, 14, 15, 16, 17, 21, and 22.

---

resulted in the conversion equations that (for each examinee) were applied to the raw scores for each task. The scaled scores for the two tasks were then summed and rounded to one decimal place to create the PT score for each candidate. The PT scores had a possible range of 0 to 13.2.

NCBE PT Questionnaire
Summary Statistics

1. Applicant Number _ _ _ _ _

2. Birthdate [/ /]
month day year

	Total	3-hour	90-minute
Mean =	31.45	31.68	31.37 years of age
SD =	7.89	8.05	7.84
Range =	23 to 91	24 to 87	23 to 91
N =	1666	400	1266

3. Gender (circle one)

		Total	3-hour	90-minute
[1]	Female	43.19 %	45.91 %	42.33 %
[2]	Male	56.81 %	54.09 %	57.67 %
		N =1681	N =403	N = 1278

4. Circle the box below that best describes your undergraduate major:

[1] Economics, Business, Accounting
[2] Physical Science, Engineering, Mathematics
[3] Natural Science, Biology, Zoology, Ecology
[4] Social Science, Anthropology, Psychology, Sociology
[5] History, Government, Political Science
[6] English, Journalism, Classical Studies, Philosophy
[7] Fine Arts
[8] Education
[9] Other

	Total	3-hour	90-minute
[1]	19.73 %	21.53 %	19.16 %
[2]	6.32 %	6.19 %	6.36 %
[3]	3.37 %	2.72 %	3.57 %
[4]	9.51 %	10.15 %	9.31 %
[5]	36.62 %	32.92 %	37.78 %
[6]	15.42 %	15.84 %	15.28 %
[7]	2.54 %	3.22 %	2.33 %
[8]	1.65 %	1.73 %	1.63 %
[9]	4.84 %	5.69 %	4.58 %
	N =1693	N =404	N =1289

5. To the best of your recollection, what was your LSAT score? (If you took the LSAT more than once, please indicate your highest score. If you did not take the LSAT or have no recollection of your approximate score, mark an X): ___

	Total	3-hour	90-minute
Mean =	37.47	37.71	37.40
SD =	5.86	6.03	5.81
Range=	13 to 55	13 to 55	17 to 52
N =	1215	285	930

6. Circle the box below that best describes your racial/ethnic background:

[1] African-American/Black, non-Hispanic
[2] American Indian/Alaskan Native
[3] Asian-American or Pacific Islander
[4] Caucasian/White, non-Hispanic
[5] Mexican-American/Chicano
[6] Puerto Rican, Cuban, other Hispanic/Latino
[7] Origins in Indian sub-continent
[8] Other

	Total	3-hour	90-minute
[1]	14.97 %	15.29 %	14.87 %
[2]	1.20 %	1.00 %	1.27 %
[3]	2.59 %	2.01 %	2.77 %
[4]	74.62 %	74.94 %	74.53 %
[5]	2.04 %	2.01 %	2.06 %
[6]	2.29 %	2.26 %	2.29 %
[7]	0.60 %	0.75 %	0.55 %
[8]	1.68 %	1.75 %	1.66 %
	N =1663	N = 399	N = 1264

7. About what percentage of the time was English spoken in your home during your childhood? Please put your answer in the box below

□ Percentage of time English spoken.

Total N = 1675

More than 85% (1427) of the candidates indicated that English was spoken in their home 100% of the time during childhood. Less than 1% (15) of the candidates reported that English was not spoken at home during childhood. Nearly 15% (248) of the candidates indicated that some language other than English was spoken at home during their childhood at least some of the time.

8. In the box next to each choice below, please indicate the number of semester credits, if any, you have earned in:

[ ] Clinical or Trial Practice courses involving your participation in simulated or actual hearings

Total N = 1718

Nearly 60% (1022) of the candidates indicated that they had not earned any semester hours in this type of clinical or trial practice course. The next largest proportion of candidate response was at three semester hours of course work (12.75%, 219). Very few of the candidates reported having more than ten semester credits in this type of clinical or trial practice course.

[ ] Lecture courses in Trial Practice NOT involving your participation in simulated or actual hearings

Total N = 1718

Nearly 25% (425) of the candidates indicated that they had not earned any semester hours in this type of lecture course. The next largest proportion of candidate response was at three semester hours of course work (18.74%, 322). Very few of the candidates reported having more than thirty semester credits in this type of lecture course.

9. **For each of the activities below, please indicate if you have done them for: SIMULATED cases, ACTUAL cases associated with supervised law school programs, and ACTUAL cases as part of paid and/or volunteer employment. Put a check in the boxes corresponding to the things you have done:**

Note: numbers in boxes represent the percentage of total who checked that box (e.g. 66 means 66% checked this box).

			ACTUAL CASES	
		SIMULATED CASES	Law School Programs	Paid or Volunteer Employment
		Condition	Condition	Condition
		3-hr 90-min.	3-hr 90-min.	3-hr 90-min.
a.	Conduct direct or cross examinations	[65] [66]	[29] [33]	[17] [22]
b.	Prepare analytical legal memoranda	[65] [64]	[49] [50]	[67] [69]
c.	Prepare briefs, petitions, motions	[67] [65]	[50] [51]	[58] [63]
d.	Present an oral argument in a legal proceeding	[73] [71]	[42] [46]	[18] [24]
				N=412 N=1253
		Total	Total	Total
		[66]	[32]	[21]
		[65]	[50]	[68]
		[65]	[51]	[62]
		[71]	[45]	[22]
				N=1665

10. **Which of the following jobs, if any, have you held for three or more months? Please check all the boxes that apply:**
Note: The numbers represent percentage of total group who checked the box.

Total N = 1402

[1] Court Reporter
[74] Law Clerk
[3] Law Enforcement Officer
[5] Legal Investigator
[5] Legal Secretary
[20] Paralegal Assistant
[36] Other law-related employment

3-hour N = 322

[0] Court Reporter
[71] Law Clerk
[2] Law Enforcement Officer
[4] Legal Investigator
[5] Legal Secretary
[23] Paralegal Assistant
[36] Other law-related employment

90-minute N = 1080

[1] Court Reporter
[74] Law Clerk
[3] Law Enforcement Officer
[5] Legal Investigator
[5] Legal Secretary
[19] Paralegal Assistant
[36] Other law-related employment

11. Have you ever been licensed to practice law? Please circle one box. If yes, please indicate the total number of years you have been licensed.

Total

[1] no — Number = 1446 (85.36% of total)
[2] yes: number of years — Number = 248 (14.64% of total)

Mean = 6.18 years
SD = 6.70 years
Range = 0 to 37
N = 242

3-hour

[1] no — Number = 351 (86.45% of total)
[2] yes: number of years — Number = 55 (13.55% of total)

Mean = 5.24 years
SD = 5.30 years
Range = 0 to 25
N = 406

90-minute

[1] no — Number = 1095 (85.02% of total)
[2] yes: number of years — Number = 193 (14.98% of total)

Mean = 6.46 years
SD = 7.04 years
Range = 0 to 37
N = 1288

12. In your opinion, how good a measure of your LEGAL REASONING ABILITY was each part of the examination? Please circle one number for each part.

	Very Poor	Poor	Fair	Good	Very Good	Did Not Take
a. Essay Examination	1	2	3	4	5	0
b. Multistate (MBE)	1	2	3	4	5	0
c. Performance Test	1	2	3	4	5	0

a. Essay Examination

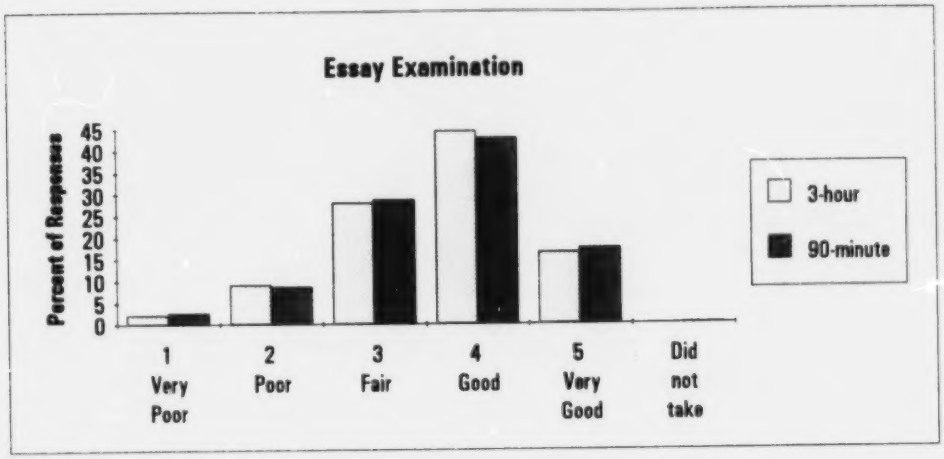
	Total	3-hour	90-minute
Mean =	3.64	3.64	3.64
SD =	0.95	0.94	0.95
N =	1695	408	1287

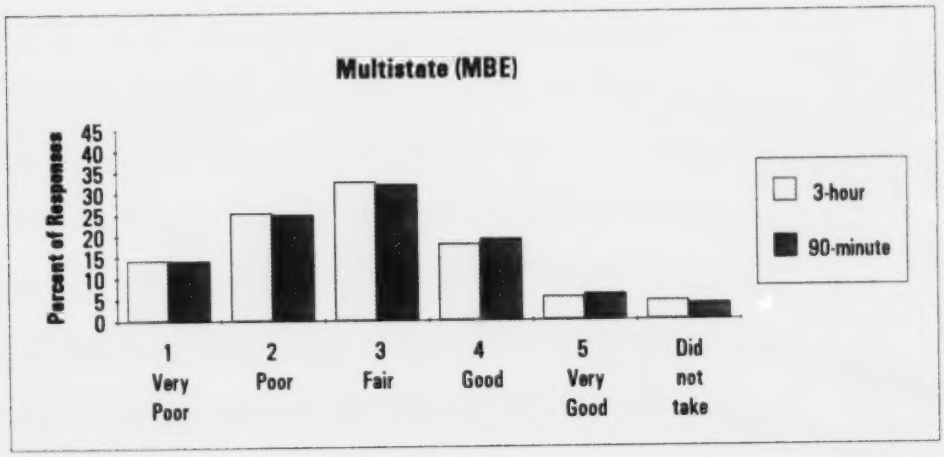
b. Multistate (MBE)

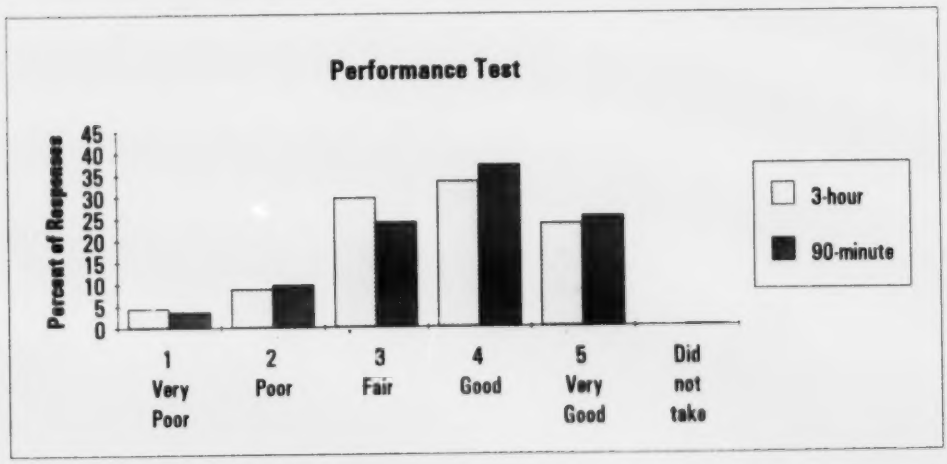
	Total	3-hour	90-minute
Mean =	2.65	2.62	2.67
SD =	1.22	1.21	1.22
N =	1687	407	1280

c. Performance Test

	Total	3-hour	90-minute
Mean =	3.68	3.63	3.70
SD =	1.07	1.07	1.07
N =	1691	407	1284


Essay Examination
Percent of Responses
45
40
35
30
25
20
15
10
5
0
1 Very Poor
2 Poor
3 Fair
4 Good
5 Very Good
Did not take
3-hour
90-minute


Multistate (MBE)
Percent of Responses
45
40
35
30
25
20
15
10
5
0
1 Very Poor
2 Poor
3 Fair
4 Good
5 Very Good
Did not take
3-hour
90-minute


Performance Test
Percent of Responses
45
40
35
30
25
20
15
10
5
0
1 Very Poor
2 Poor
3 Fair
4 Good
5 Very Good
Did not take
3-hour
90-minute

13. In your opinion, how well did each examination cover your FULL RANGE OF LEGAL KNOWLEDGE? Please circle one number for each part.

	Very Poor	Poor	Fair	Good	Very Good	Did Not Take
a. Essay Examination	1	2	3	4	5	0
b. Multistate (MBE)	1	2	3	4	5	0
c. Performance Test	1	2	3	4	5	0

a. Essay Examination

	Total	3-hour	90-minute
Mean =	3.36	3.37	3.35
SD =	1.09	1.12	1.08
N =	1695	405	1290

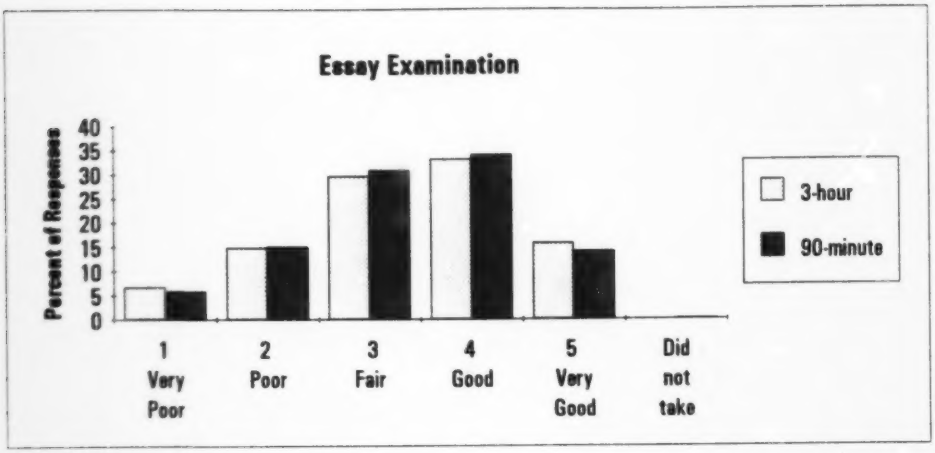
b. Multistate (MBE)

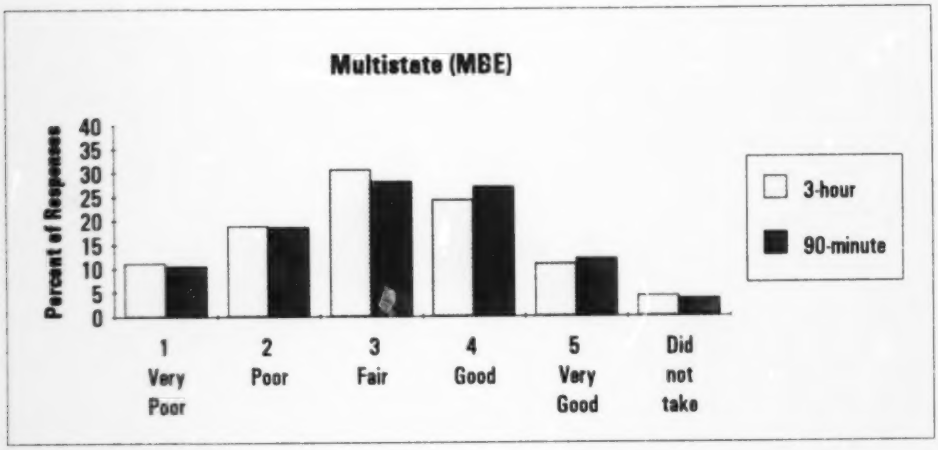
	Total	3-hour	90-minute
Mean =	2.98	2.92	3.00
SD =	1.30	1.30	1.30
N =	1686	404	1282

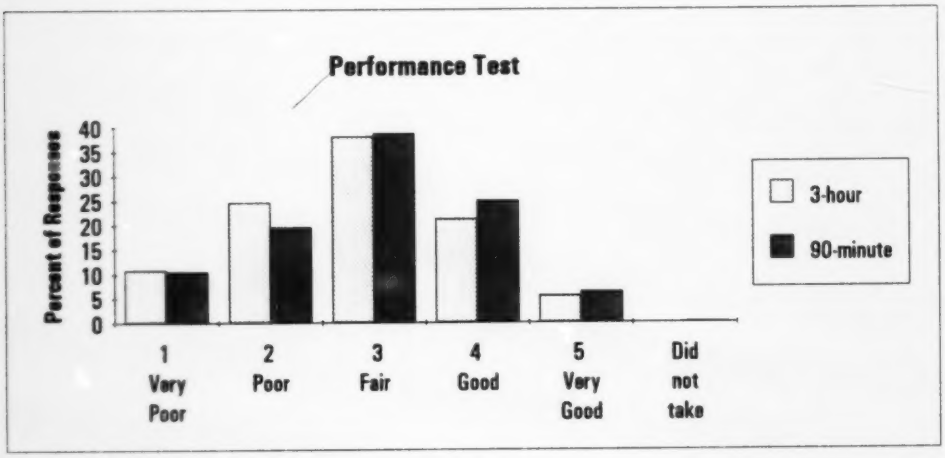
c. Performance Test

	Total	3-hour	90-minute
Mean =	2.94	2.86	2.97
SD =	1.06	1.04	1.06
N =	1692	407	1285

Examination Covered Full Range of Legal Knowledge







14. In your opinion, how good a measure of your ABILITY TO PERFORM AS AN ATTORNEY was each part of the examination? Please circle one number for each part.

	Very Poor	Poor	Fair	Good	Very Good	Did Not Take
a. Essay Examination	1	2	3	4	5	0
b. Multistate (MBE)	1	2	3	4	5	0
c. Performance Test	1	2	3	4	5	0

a. Essay Examination

	Total	3-hour	90-minute
Mean =	2.96	2.90	2.98
SD =	1.14	1.14	1.14
N =	1686	403	1283

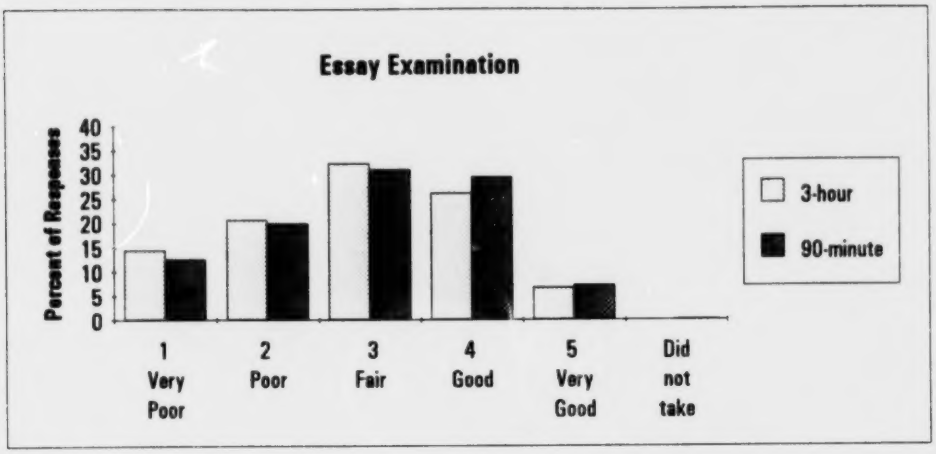
b. Multistate (MBE)

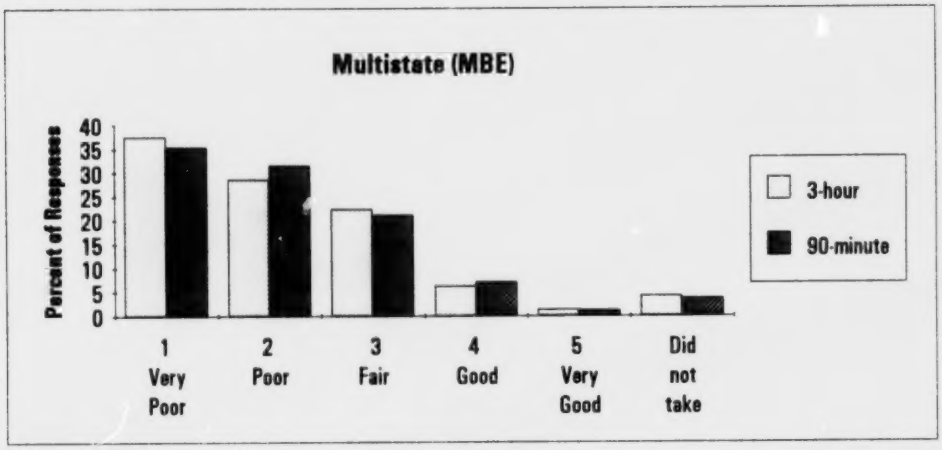
	Total	3-hour	90-minute
Mean =	1.95	1.93	1.96
SD =	1.05	1.06	1.05
N =	1680	403	1277

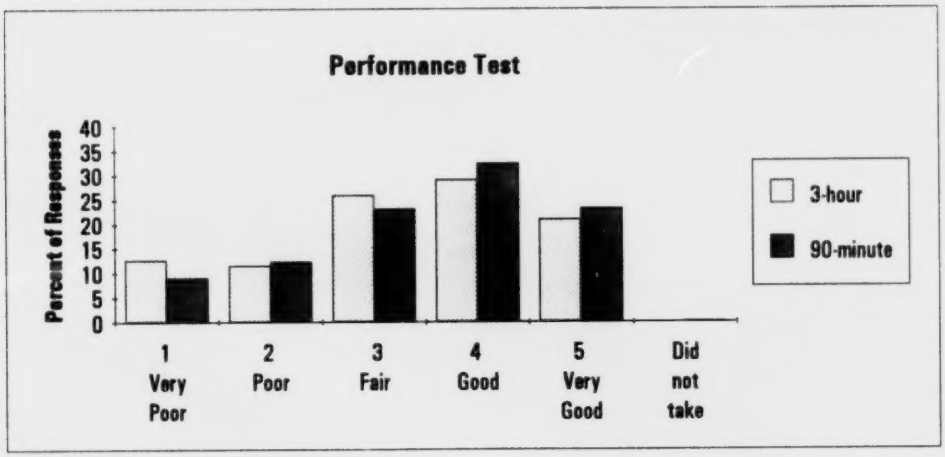
c. Performance Test

	Total	3-hour	90-minute
Mean =	3.45	3.34	3.48
SD =	1.24	1.28	1.23
N =	1680	402	1278

Measure of Ability to Perform as an Attorney







15. In your opinion, HOW DIFFICULT WAS EACH PART of the examination? Please circle one number for each part.

	Very Difficult	Difficult	Average	Easy	Very Easy	Did Not Take
a. Essay Examination	1	2	3	4	5	0
b. Multistate (MBE)	1	2	3	4	5	0
c. Performance Test	1	2	3	4	5	0

a. Essay Examination

	Total	3-hour	90-minute
Mean =	2.33	2.33	2.33
SD =	0.77	0.80	0.76
N =	1687	401	1286

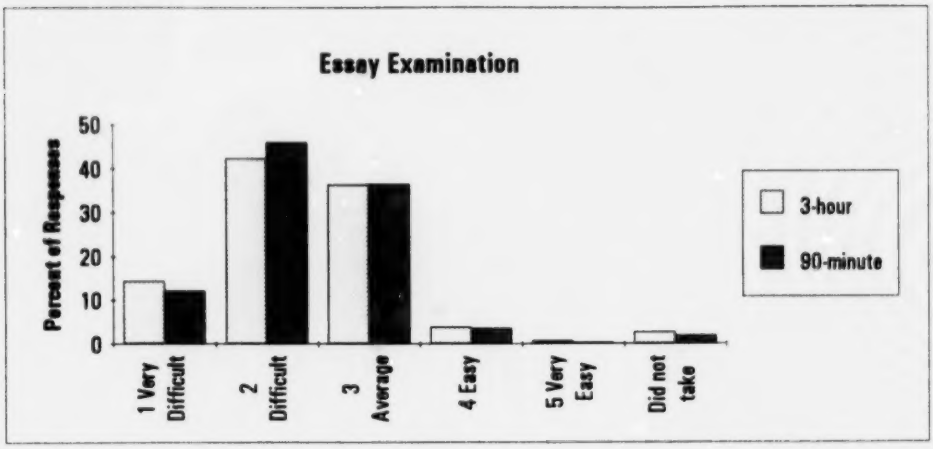
b. Multistate (MBE)

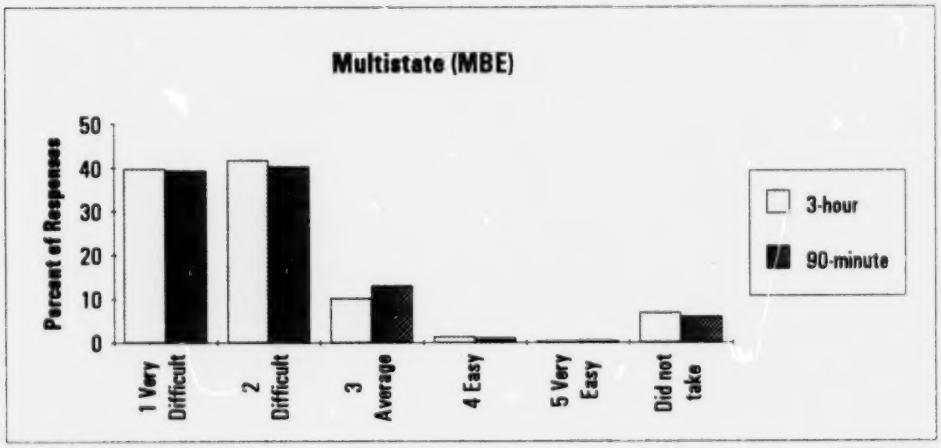
	Total	3-hour	90-minute
Mean =	1.67	1.64	1.69
SD =	0.82	0.80	0.82
N =	1682	402	1280

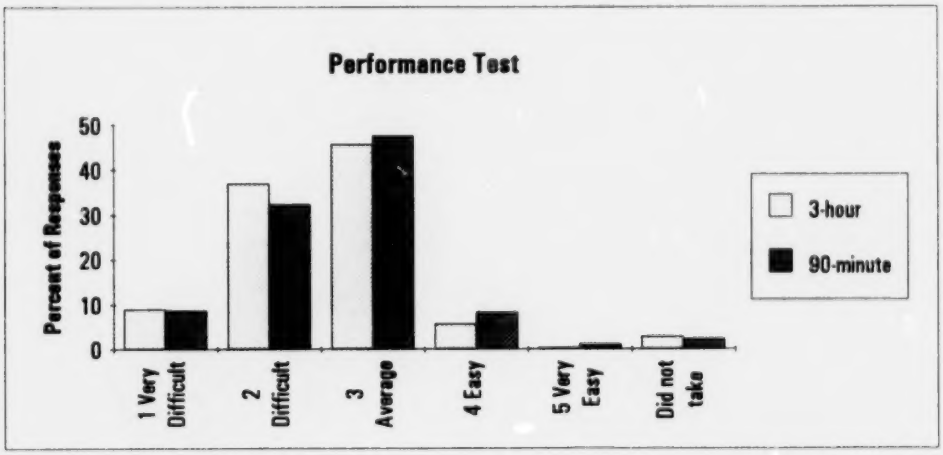
c. Performance Test

	Total	3-hour	90-minute
Mean =	2.57	2.50	2.60
SD =	0.80	0.75	0.80
N =	1678	401	1277

Difficulty of Each Examination







16. In your opinion, WAS THE TIME ALLOWED FOR EACH PART OF THE EXAMINATION APPROPRIATE? Please circle one number for each part.

	Much Less Than Enough	Less Than Enough	About Right	More Than Enough	Much More Than Enough	Did Not Take
a. Essay Examination	1	2	3	4	5	0
b. Multistate (MBE)	1	2	3	4	5	0
c. Performance Test	1	2	3	4	5	0

a. Essay Examination

	Total	3-hour	90-minute
Mean =	2.84	2.81	2.85
SD =	0.58	0.62	0.57
N =	1684	402	1282

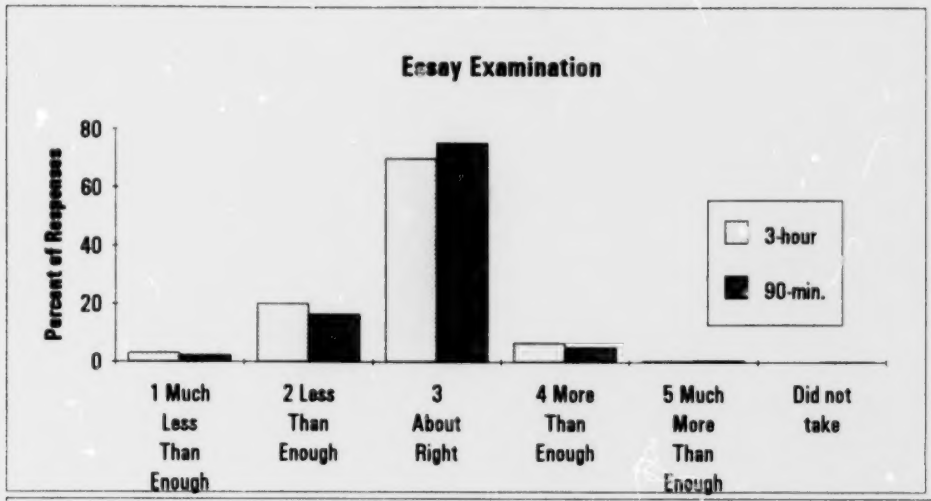
b. Multistate (MBE)

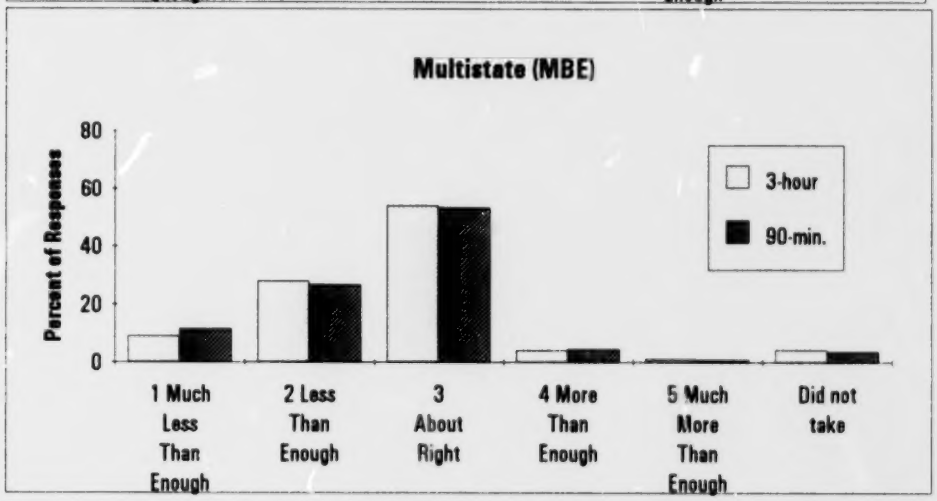
	Total	3-hour	90-minute
Mean =	2.46	2.48	2.46
SD =	0.91	0.91	0.91
N =	1674	402	1272

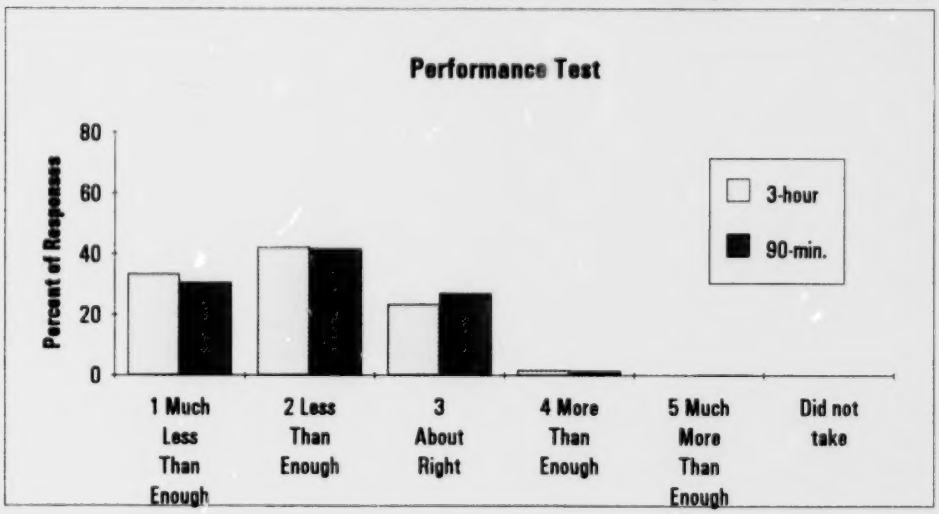
c. Performance Test

	Total	3-hour	90-minute
Mean =	1.98	1.93	1.99
SD =	0.80	0.79	0.80
N =	1686	402	1284

Time Allowed for Each Part







17. If you were responsible for licensing attorneys in this state, HOW HEAVILY WOULD YOU WEIGHT EACH OF THE TESTS in making the licensure decision? Assume that all other eligibility requirements, such as screening for character and fitness, have been met. Place a percent value by each of the parts so that the total for all parts equals 100%.

a. Essay

	Mean	SD	N
Total	42.62 %	17.77	1718
3-hour	43.36 %	18.09	412
90-minute	42.39 %	17.67	1306

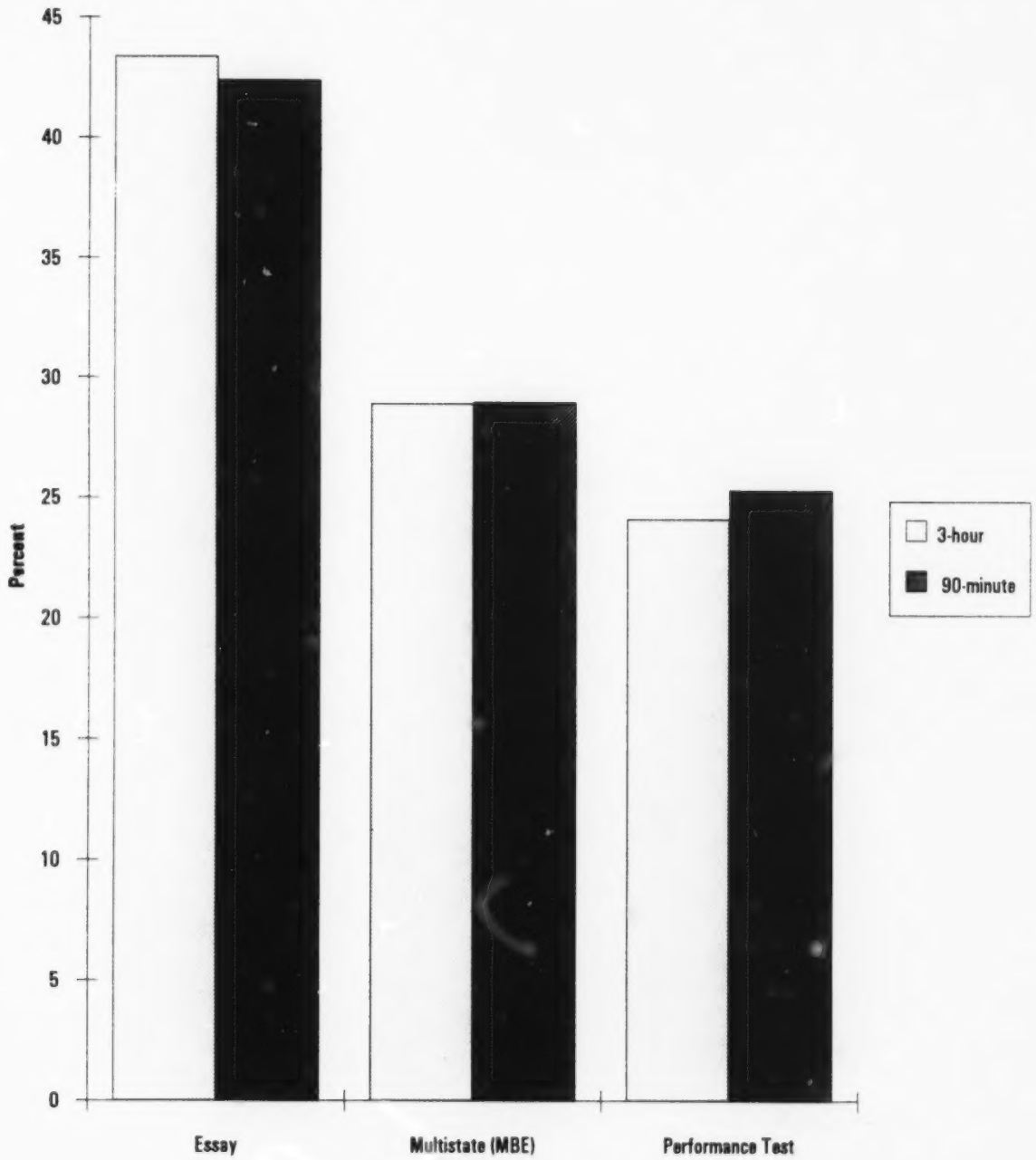
b. Multistate (MBE)

	Mean	SD	N
Total	28.95 %	15.52	1718
3-hour	28.91 %	16.03	412
90-minute	28.96 %	15.36	1306

c. Performance Test

	Mean	SD	N
Total	25.02 %	16.42	1708
3-hour	24.10 %	16.26	412
90-minute	25.30 %	16.47	1306

How Would You Weigh Each Examination in Making Licensure Decisions?



18. Circle one number for each performance test to indicate the ADEQUACY OF THE FILE MATERIALS for preparing your written answer:

		Not Adequate				Very Adequate	Did Not Take
a.	Dodson -- cross-examination plan	1	2	3	4	5	0
b.	Dodson -- analytical memo	1	2	3	4	5	0
c.	Xenophanes -- persuasive memo	1	2	3	4	5	0
d.	Xenophanes -- closing argument	1	2	3	4	5	0

a. Dodson -- plan

	Mean	SD	N
Total	2.77	1.88	1395
3-hour	2.70	1.82	335
90-minute	2.79	1.90	1060

b. Dodson -- memo

	Mean	SD	N
Total	2.60	1.85	1389
3-hour	2.48	1.86	327
90-minute	2.64	1.85	1062

c. Xenophanes -- memo

	Mean	SD	N
Total	2.76	1.89	1413
3-hour	2.70	1.90	343
90-minute	2.78	1.89	1070

d. Xenophanes -- close

	Mean	SD	N
Total	2.78	1.96	1408
3-hour	2.81	1.91	343
90-minute	2.76	1.98	1065

19. Circle one number for each performance test to indicate the ADEQUACY OF THE LIBRARY MATERIALS for preparing your written answer:

		Not Adequate				Very Adequate	Did Not Take
a.	Dodson -- cross-examination plan	1	2	3	4	5	0
b.	Dodson -- analytical memo	1	2	3	4	5	0
c.	Xenophanes -- persuasive memo	1	2	3	4	5	0
d.	Xenophanes -- closing argument	1	2	3	4	5	0

a. Dodson -- plan

	Mean	SD	N
Total	2.72	1.85	1375
3-hour	2.73	1.80	326
90-minute	2.71	1.86	1049

b. Dodson -- memo

	Mean	SD	N
Total	2.59	1.82	1372
3-hour	2.50	1.84	322
90-minute	2.62	1.82	1050

c. Xenophanes -- memo

	Mean	SD	N
Total	2.71	1.86	1386
3-hour	2.65	1.87	331
90-minute	2.74	1.86	1055

d. Xenophanes -- close

	Mean	SD	N
Total	2.79	1.94	1382
3-hour	2.82	1.90	334
90-minute	2.78	1.96	1048

20. Circle one number for each performance test to indicate the ADEQUACY OF ITS INSTRUCTIONS:

		Not Adequate				Very Adequate	Did Not Take
a.	Dodson -- cross-examination plan	1	2	3	4	5	0
b.	Dodson -- analytical memo	1	2	3	4	5	0
c.	Xenophanes -- persuasive memo	1	2	3	4	5	0
d.	Xenophanes -- closing argument	1	2	3	4	5	0

a. Dodson -- plan

	Mean	SD	N
Total	2.96	1.96	1351
3-hour	3.00	1.93	320
90-minute	2.95	1.97	1031

b. Dodson -- memo

	Mean	SD	N
Total	2.71	1.91	1345
3-hour	2.62	1.89	318
90-minute	2.74	1.91	1027

c. Xenophanes -- memo

	Mean	SD	N
Total	2.94	1.96	1365
3-hour	2.84	1.98	330
90-minute	2.96	1.95	1035

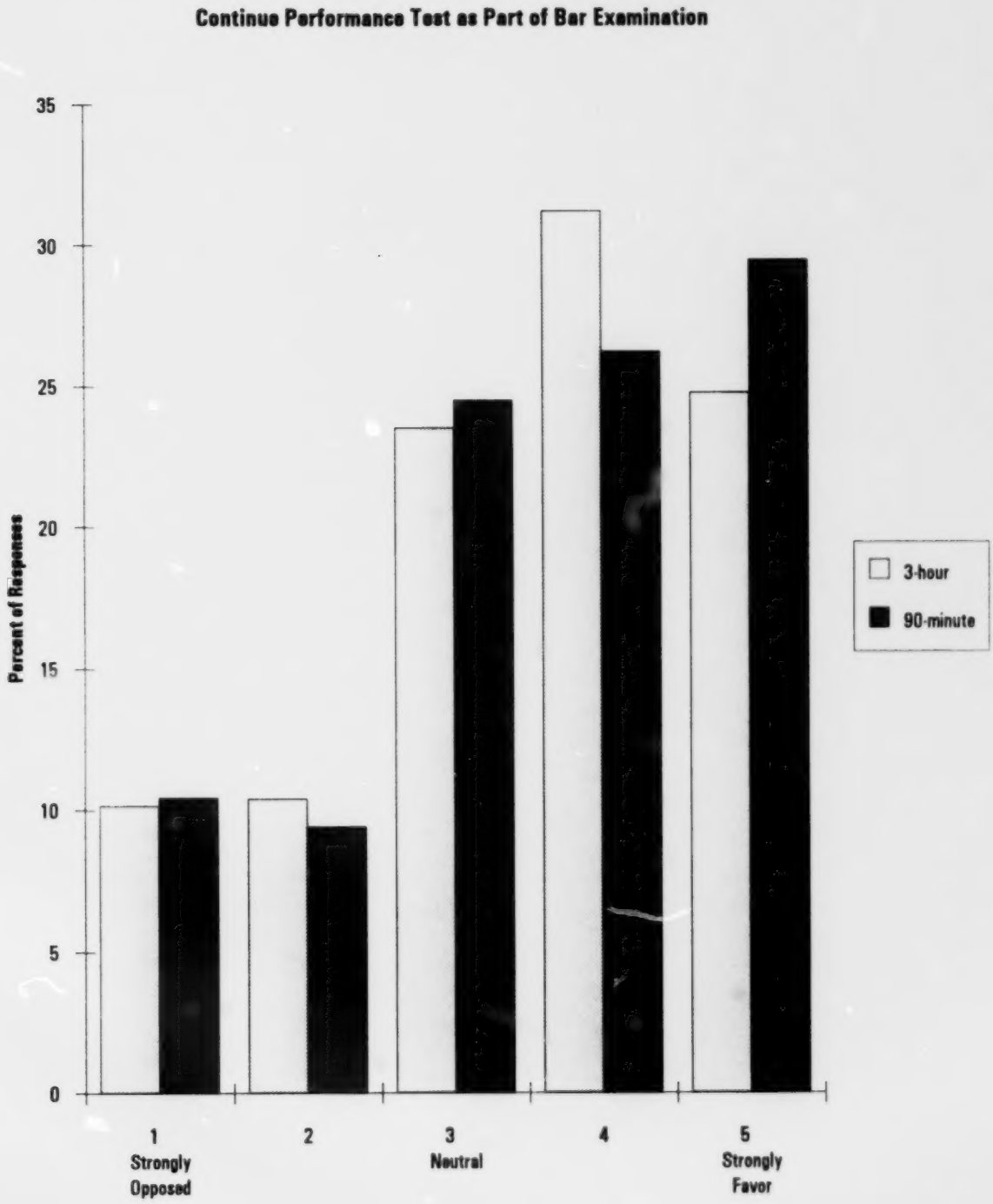
d. Xenophanes -- close

	Mean	SD	N
Total	2.92	2.00	1365
3-hour	2.92	1.97	333
90-minute	2.93	2.02	1032

21. What is your opinion regarding whether the performance test should continue to be a regular part of the bar exam? Circle one number:

Strongly Opposed		Neutral		Strongly Favor
1	2	3	4	5

	Mean	SD	N
Total	3.54	1.28	1670
3-hour	3.50	1.25	404
90-minute	3.55	1.29	1266


Continue Performance Test as Part of Bar Examination
Percent of Responses
35
30
25
20
15
10
5
0
1
Strongly
Opposed
2
3
Neutral
4
5
Strongly
Favor
3-hour
90-minute

Item 22 Response Summary

Following is a brief summary of bar candidates' narrative reactions to the NCBE Performance Test.

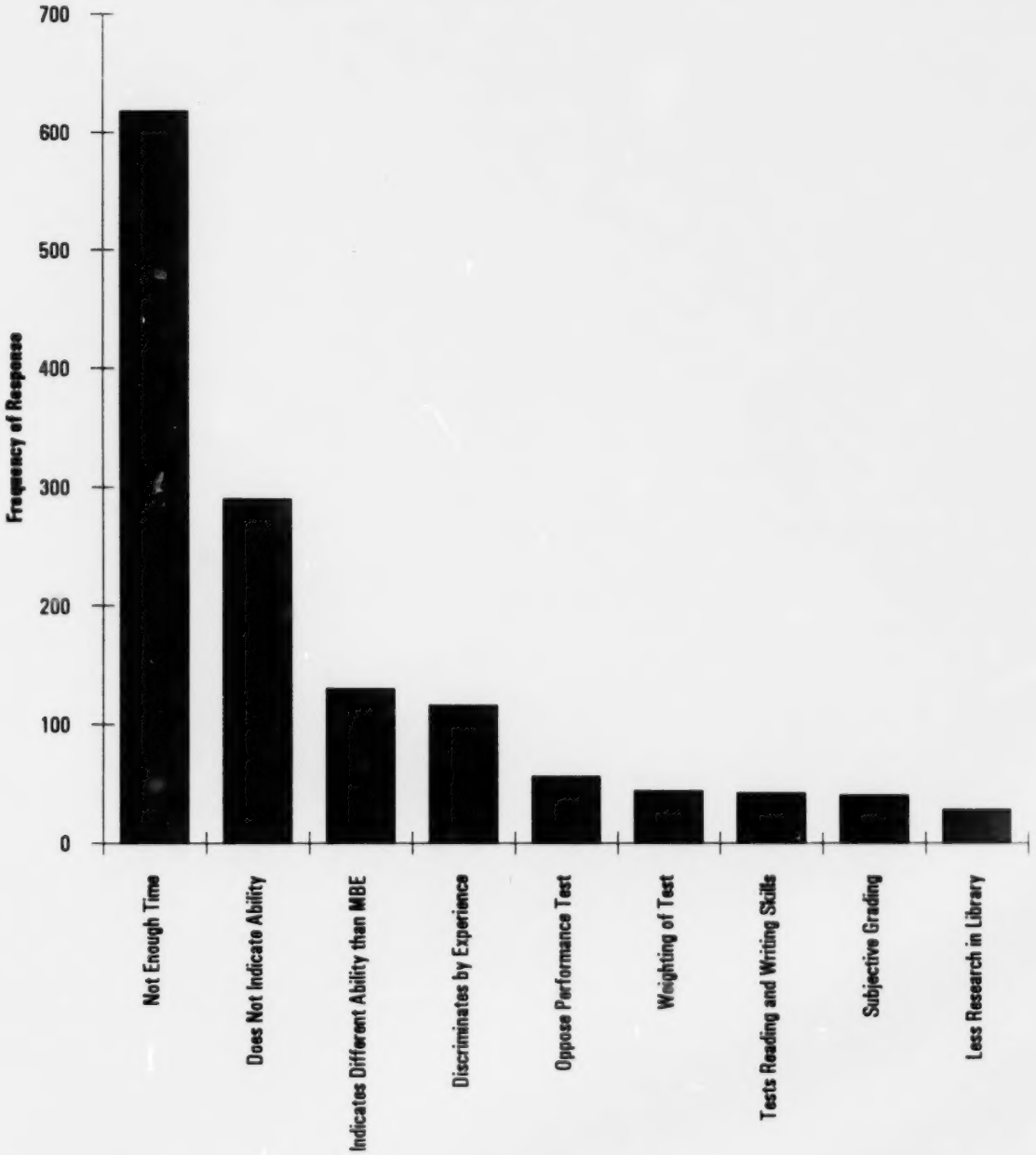
Most of the comments dealt with the issue of time. Respondents raising this issue thought that there simply was not enough time to thoroughly digest the materials and prepare a response that would accurately reflect a person's ability. Here are a few comments on this topic.

> "It [the Performance Test] could be improved by eliminating the time requirement."
>
> "It [the lack of time] encourages malpractice."
>
> "Do you want to encourage future lawyers to skim opinions without digesting them for their clients?"
>
> "I certainly would have appreciated at least two more hours to take this test."
>
> "Shorten Xenophanes or lengthen the time by 90 minutes."

A significant number of candidates thought the Performance Test was biased against inexperienced lawyers: ("It would be great if this kind of hands-on performance was actually taught in law schools, therefore to be fair we need to teach performance skills." -- "I think the value of the Performance Test in determining an applicant's qualifications for the practice of law is substantially outweighed by its prejudice against recent law graduates.") Others saw the Performance Test as biased against lawyers without litigation experience: ("I am concerned that a student with no trial advocacy background could not write the closing argument because of lack of background.")

Overall, most of the comments were negative in nature, and most of these had to do with the issue of insufficient time. The other major subjects addressed are shown on the graph for item 22 (following).

Topics Addressed in Item 22:
General Reactions to Performance Test



Relationship Analyses

As described earlier, total PT scaled scores were correlated with a number of variables from the questionnaire. Each of these correlations was then tested to see if it was significantly different from zero. The results, with comments, are summarized in the following table.

Variable	PT Question Number	Correlation with Total PT Scaled Score	Significant at α=.01 (*)	Comment
Age	2	.00		
Gender	3	.00		
Reported LSAT Score	5	.02		
% English spoken	7	-.03		
Credits in Practice courses	8	.03		
Credits in Lecture courses	8	.01		
Experience variables (1 of 12 was significant):	9			
Prepare analytical legal memoranda on actual cases, paid or volunteer employment	9b (3)	-.07	*	Candidates with this experience had slight tendency to score lower on PT.
Employment history	10	-.03 to +.04		Seven jobs were listed. None were significantly correlated with PT score.
Licensed to practice	11	.06		
Years licensed	11	-.04		
Opinion of Essay as measure of legal reasoning	12	-.03		
Opinion of MBE as measure of legal reasoning	12	-.04		

Variable	PT Question Number	Correlation with Total PT Scaled Score	Significant at α=.01 (*)	Comment
Opinion of PT as measure of legal reasoning	12	-.05		
Opinion of Essay covering full range of knowledge	13	.01		
Opinion of MBE covering full range of knowledge	13	-.02		
Opinion of PT covering full range of knowledge	13	-.03		
Opinion of Essay as measure of ability to perform	14	-.05		
Opinion of MBE as measure of ability to perform	14	-.03		
Opinion of PT as measure of ability to perform	14	-.04		
Opinion of Essay difficulty	15	-.07	*	Candidates who rated essay as easier had slight tendency to score lower on the PT.
Opinion of MBE difficulty	15	-.03		
Opinion of PT difficulty	15	-.04		
Opinion of Essay time sufficiency	16	-.01		
Opinion of MBE time sufficiency	16	-.04		
Opinion of PT time sufficiency	16	-.01		
Opinion of Essay weighting in licensure	17	.04		
Opinion of MBE weighting in licensure	17	-.03		

Variable	PT Question Number	Correlation with Total PT Scaled Score	Significant at α=.01 (*)	Comment
Opinion of PT weighting in licensure	17	-.03		
Adequacy of file materials (4 items)	18	-.03 to +.03		No significant relationships.
Adequacy of library materials (4 items)	19	-.03 to +.01		No significant relationships.
Adequacy of instructions (4 items)	20	-.03 to +.01		No significant relationships.
Opinion on continuing PT	21	-.04		

Discussion

Response Tabulations

Perhaps one of the most striking aspects of the results is the similarity of the characteristics of the 3-hour and 90-minute groups. The groups were nearly identical in every significant demographic and attitudinal variable in the questionnaire. The demographic similarity is an indicator of the equivalence of the groups. The random assignment of candidates appears to have resulted in this equivalence.

Beginning with question 2, we see that the mean age was between 31 and 32 years. Because the range goes up to 91, however, we suspect that some examinees may have misstated their birth year. If so, the actual mean age of the group would be a bit lower. The other demographic and background information in items 3 through 11 is self-explanatory.

Opinion items begin with number 12. On that item, examinees viewed the Essay and PT exams about equally as "fair" to "good" measures of legal reasoning ability. The mean rating of the MBE was about one scale point lower than the other two, placing it in the "poor" to "fair" range. Bar graphs are provided to display the distributions of ratings.

Opinions of the three exams as measures of the full range of legal knowledge were assessed in question 13. The Essay was rated somewhat higher than the others with a mean rating of about 3.36, which is a bit better than "fair." The MBE and PT both scored near the scale midpoint with 2.98 and 2.94, respectively.

A clear rank ordering of the three exams was evident in number 14, the rating of tests as measures of ability to perform as an attorney. The PT was judged "fair" to "good" in this regard, the Essay "fair," and the MBE "poor." Of course, these are point descriptors of group mean ratings, and there was considerable spread (disagreement) within the groups.

Item 15 addressed the perceived difficulty of the exams. The MBE was judged to be clearly the most difficult, with over 80% of examinees rating it as "difficult" or "very difficult." By contrast, the ratings of the Essay and PT exams were in the "difficult" to "average" range. This result may provide a clue as to some of the other perceptions of the MBE. Examinees who perceived the MBE to be too difficult for them might not be inclined to rate it highly in other respects.

The results in item 16 seem very relevant to future PT development. About 75% of all candidates felt that they had either "less than enough" or "much less than enough time" to complete the PT. Interestingly, this was true of both the 3-hour and 90-minute groups. In fact, the ratings of the two groups were nearly identical. Most of the ratings of the Essay exam indicated that its timing was "about right," while the

MBE ratings were between the other two. This item should serve as a flag that the speededness issue should be examined more closely for the PT.

Item 17 asked candidates to assign ideal relative weights to the three exams in a licensure context. This item was designed to challenge candidates to make a "bottom line" evaluation of the relative value of the three examinations. Again, the 3-hour and 90-minute conditions' ratings were very similar. The Essay exam would receive the heaviest weights, followed by the MBE, with the least weight to the PT. The largest difference was between the Essay and the other two. It is interesting to consider these perceptions in comparison to the other ratings of test difficulty, speededness, and face validity (legal reasoning, ability to perform as an attorney, etc.). Another relevant factor is the total time available for the three tests; i.e., six hours each for the essay and MBE and three hours for the PT. Considering all these factors, one might speculate as to the reasons for the results in relative weightings. One possibility is that the Essay exam was rated highly because the format is very familiar to law students, they felt they had enough time for the test, and it didn't seem too difficult. The MBE, on the other hand, was perceived as very difficult, and perhaps was rated lower as a result. Likewise, the speededness issue may have resulted in lower ratings of the PT.

Items 18, 19, and 20 were designed to provide guidance to writers of any future forms of the PT. Most of the mean ratings of the PT materials and instructions were just below the scale midpoints.

Item 21 called for another "bottom line" judgment on the part of examinees regarding whether the PT should be continued. The mean rating was about 3.54 on a scale of 1 to 5, with 3 being neutral and 5 being strongly in favor. Approximately 55% of the candidates favored its continuation, about 25% were neutral, and about 20% were opposed.

Item 22 responses were summarized earlier. The free-form response format of this item may have elicited more negative comments about the PT.

Relationship Analyses

Clearly, a tremendous amount of data were obtained from this questionnaire, creating almost limitless possibilities for data exploration. Because our immediate objective concerned performance on the PT, we decided to limit this phase of the analysis to correlations of questionnaire variables with PT scores. The PT scores used were total scaled scores. This provided a common metric so that the various candidate groups who took different PT tasks in different time limits would be combined into one "total" candidate population.

Interestingly, both age and gender correlated virtually zero with PT performance. The age result is difficult to interpret because an experience factor is likely to be

confounded in this. The zero gender correlation provides evidence of no significant differential performance between the sexes.

One might expect a moderate positive correlation of PT score with LSAT score as an indicator that there is some commonality between what the instruments are measuring. However, the correlation of +.02 is not significantly different from zero. It is possible that actual LSAT scores as opposed to self-reported ones would correlate more highly.

The question about the percentage of time that English was spoken in the home gives us the opportunity to detect differential performance by individuals who may have English as a second language. The nonsignificant correlation might be construed to mean that no such concerns are raised by these data. However, the highly restricted range should serve as a caution against overinterpretation of this result.

Questions 8, 9, and 10 provided extensive information on the preparation of candidates in terms of courses taken, practice experiences, and employment. None of the correlations between PT scores and these variables were significantly different from zero. The test is apparently measuring something unrelated to these variables. However, it should be noted that these are self-reported experiences, and actual documented experience variables may correlate differently.

The reported presence of a license to practice law and the number of years licensed were not related to PT score. Again, the PT seems to be measuring something unrelated to these variables.

The remainder of the correlational studies concerned the relationships of total PT scaled scores with various opinions. For the most part these were near zero and unremarkable, except that it is interesting that candidates who rated the Essay as easier had a slight tendency to score lower on the PT.

In summary, it appears that the PT is a concept that makes intuitive sense to many law candidates. They see it as a viable method of assessing lawyering ability. However, they apparently do not see it as a substitute or replacement for the other examinations. A fictitious "average candidate" might say that each of the exams has its place in the licensure process, or at least that none is clearly better or worse than the others. The Essay exams are fairly well liked as they are. The MBE is seen as very difficult and the PT is seen as being too much work for the time allowed. The relationship analyses involved data exploration to determine if the PT was measuring any constructs in common with the demographic, experience, or attitude items in the questionnaire. These analyses provided little if any evidence of such commonality.